Utilizing Grasp Monitoring to Predict Microsurgical Expertise

Jani Koskinen, MSc,a,* Wenjing He, PhD,b Antti-Pekka Elomaa, PhD,c,d,e Aku Kaipainen, MD,c,d Ahmed Hussein, MD,c,d Bin Zheng, PhD,f Antti Huotarinen, PhD,c,d and Roman Bednarik, PhDa

a School of Computing, University of Eastern Finland, Joensuu, Finland
b Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
c Department of Neurosurgery, Institute of Clinical Medicine, Kuopio University Hospital, Kuopio, Finland
d Microsurgery Center, Kuopio University Hospital, Kuopio, Finland
e School of Medicine, University of Eastern Finland, Kuopio, Finland
f Surgical Simulation Research Lab, Department of Surgery, University of Alberta, Edmonton, Alberta, Canada

A B S T R A C T

Introduction: Most microsurgical procedures require the surgeon to use tools to grasp and hold fragile objects in the surgical site. Prior research on grasping in surgery has mostly either been in other surgical techniques or used grasping as an auxiliary metric. We focus on microsurgery and investigate what grasping can tell about microsurgical skill and suturing performance. This study lays groundwork for using automatic detection of grasps to evaluate surgical skill.

Methods: Five expert surgeons and six novices completed sutures on a microsurgical training board. Video recordings of the performance were annotated for the number of grasps, while an eye tracker recorded the participants’ pupil dilations for cognitive workload assessment. Performance was measured with suturing duration and the University of Western Ontario Microsurgical Skills Assessment instrument (UWOMSA). Differences in skill, suturing performance and cognitive workload were compared with grasping behavior.

Results: Novices needed significantly more grasps to complete sutures and failed to grasp more often than the experts. The number of grasps affected the suturing duration more in novices. Decreasing suturing efficiency as measured by UWOMSA instrument was associated with increase in grasps, even when we controlled for overall skill differences. Novices displayed larger pupil dilations when averaged over a sufficiently large sample, and the difference increased after the grasp.

Conclusions: Grasping action during microsurgical procedures can be used as a conceptually simple yet objective proxy in microsurgical performance assessment. If the grasps could be detected automatically, they could be used to aid in computational evaluation of surgical trainees’ performance.

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Introduction

In many surgical fields, video images or a surgical microscope mediate the tool movements to the surgeon. To control tools in such settings, surgeons will need time to practice. When they have gained enough experience their handling of the tools will become fluent, indicating a cognitive merge between the tool and human motor system.\(^1\) After this point, the surgeons can focus on using the tools despite distractions.\(^5\) In contrast, inexperienced surgeons’ minds are consumed by tool control,\(^4\) and their manipulation of the surgical task will become awkward and interrupted.\(^6\) Analysis of tool use could thus reveal surgical skill.

Here we focus on grasping and examine how it is correlated with microsurgical skill, with the aim of assessing whether it would be feasible to use automatic detection of grasps to evaluate microsurgical skill in the future. Earlier research on grasping in surgery has focused mainly on measuring grasping forces.\(^10,11\) Kazemi et al. used the number of grasps needed to complete a task to measure surgical performance.\(^12\) Some studies have investigated how grasps increase task duration.\(^13\) Failure to grasp has been used in evaluating surgical errors,\(^9,16\) and Law et al. found that surgical residents display grasping difficulties when practicing with laparoscopic tools.\(^12\) These prior studies, however, have mainly considered grasping as an auxiliary metric. We expect that grasping is a fundamental yet complex action with potential to reveal crucial differences in microsurgical expertise.

Specifically, in this project, we are interested in grasping behavior during microsurgical suturing. To complete a suture, the surgeon must use tools to repeatedly grasp suture material and fragile tissues in confined spaces under the microscope while carefully applying force to prevent damage or dropping of objects. These challenges may affect expert and novice surgeons differently. Expert surgeons, for whom grasping with tools is not challenging, can spend more mental resources on other aspects of surgical performance and decision-making.\(^30\) Novices, on the other hand, may struggle with basic manipulations, like grasping with tools, to the level of frustration\(^14\) — ultimately leading to inferior sutures.

Prior research has highlighted the need for developing more objective surgical skill assessment methods.\(^27,28\) We believe that grasp monitoring could be used in objective assessment of surgical performance, similarly to the analyses of tool movements that have been conducted in previous studies.\(^29\) Specifically, we hypothesize that: 1) novices will display a greater number of grasps than experts during suturing; 2) additional grasps will increase the task time more with novices; 3) sutures that required more grasps have a lower surgical performance score; 4) the novices will have a higher level of cognitive load during grasping as measured by increasing pupil dilation.

Methods

Participants

The study was conducted at the Surgical Simulation Research Laboratory at the University of Alberta, Canada. Eleven participants were recruited, and all gave a written consent to participate in the study. The background of the participants is summarized in Table 1. The table shows the participants’ prior experience microsurgical techniques and in other surgical fields, measured in months of professional experience. The experts came from a plastic surgery clinic and performed 30 — 60 microsurgical procedures per month with either a surgical microscope or magnifying loupes. The novices were recruited from the surgical simulation laboratory where the data was recorded. They were familiar with the suturing task but did not work professionally as surgeons. They were designated as novices by their lack of prior experience in microsurgical techniques. The study was approved by a Health Research Ethics Review Board at the University of Alberta (Ethics approval number Pro00075995) and conducted following the declaration of Helsinki.

Experimental design

The experiment required the participants to complete sutures on a microsurgical training board. The board had six blocks each with a latex skin that had a pre-cut incision in the middle, and the incisions were aligned at 0, 45, or 90-degree angles. Figure 1 displays the experimental setup and the major phases of the suturing task.

The participants completed two sutures to each block, 12 total, and the first six and the last six sutures were done with different magnifications. The sutures were done with microforces, needle holder and micro-scissors, and an unused suturing needle (7/0 Prolene Blue 1 x 3” CTC-6L Double Armed, length 13 mm) with a thread (length 8 cm). The surgical microscope was Zeiss OPMI50 Vario S88 and it was equipped with a custom-made eye-tracker to record the participant’s pupil.\(^32,33\) The eye tracker was placed at the right ocular and recorded the participant’s eye at 16 – 28 Hz. Each participant’s pupil data was post-processed by up sampling it to 30 Hz and then low pass filtering at 4 Hz cutoff rate to remove noise. The operating field under the microscope was recorded for the video analysis.

<table>
<thead>
<tr>
<th>Group</th>
<th>Gender</th>
<th>Age [years]</th>
<th>Surgical experience [months]</th>
<th>Microsurgical experience [months]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Novices</td>
<td>4</td>
<td>2</td>
<td>30.5 (8.3)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>Experts</td>
<td>5</td>
<td>0</td>
<td>31.4 (1.4)</td>
<td>61.2 (19.0)</td>
</tr>
</tbody>
</table>
Assessment of suturing performance

From the full video recordings, we extracted clips containing 2 sutures. The video clips were first anonymized and then viewed in random order by an experienced neurosurgeon, who assessed the suturing performance using the University of Western Ontario Microsurgical Skills Acquisition Instrument (UWOMSA) instrument. The UWOMSA instrument is a grading scale that expert surgeons can use to systematically evaluate trainee’s performance in microsurgical knot-tying and anastomosis tasks. The instrument has separate modules for the two tasks where performance is evaluated on a 5-point Likert scale with three different items. In the knot-tying module, the three items were: (A) quality of the knot (square knot, snugly tied, ends are cut at proper length); (B) efficiency (number of wasted movements, number of grasps, needle is pulled out of field), and (C) handling (number of passes, needle is grasped at the correct location, bolsters, needle is pulled out on the curve).

A randomly chosen sample of 19 sutures were assessed by two other neurosurgeons to estimate inter-rater validity using Kendall’s rank correlation.

Video analysis of grasps

Due to the time required by manual annotation, we analyzed only the first six sutures from each participant. We annotated all the attempts of grasping the needle or the thread. Other infrequent grasp types were ignored, such as if the participant grasped the latex skin. The annotation was done by marking down the frame number and the grasp result (successful/failed grasp). Grasps were marked by the moment when the instrument tip gripped the target. Failed grasps were determined as described in Figure 2. We considered the grasp successful even if the grasped object slipped during later transportation. A video showing the suturing task with the annotated grasps is included in the Supplementary Materials.
Pupil response analysis

The pupil dilates in response to increased cognitive workload. We studied the pupil dilations in 4 s windows centered around the grasp. The pupil size changes were measured as percentage change from the first frame of each window. We removed the grasps that were within 4 s of the previous grasp, or where the pupil increased over 20%, because an increase this high would unlikely be due to increase in cognitive workload. We compared the mean pupil size change and the distance between minimum and maximum pupil sizes before and after the grasp between novices and experts.

The pupil windows around the grasps were analyzed in two ways. First, we computed an average pupil window from all grasps within individual single sutures, resulting in a total of 60 averaged pupil windows. Second, we computed an average pupil window for novices and experts, resulting in 2 pupil windows. To get the standard deviations of the mean and min-max values for the second approach we used bootstrapping by sampling with replacement novice (n = 276) and expert (n = 172) pupil windows, calculating the mean and min-max values. The process was repeated 10,000 times and the standard deviations were the standard deviation of these 10,000 values.

Statistical analyses

The effect of grasps was modeled using linear mixed effects models with participant as the random effect to account for repeated measurements from each participant. We also investigated the odds of failure to grasp with mixed effects logistic regression and t-tests in the second level of pupil analysis when pupil dilations were averaged by skill. Model fit was assessed visually with Q-Q plots and residual plots. Based on the assessment, we adjusted the models by for example, loosening the assumptions for the variance-covariance structure—ultimately resulting in an approvable fit.

To evaluate the validity of the results, we used leave-one-out cross-validation (LOOC). In LOOC, we leave out one data point at a time and fit the model with the remaining data, and then use the fitted model to predict the left-out data point. We then compare the predicted values to the actual values to get the prediction root mean sum of squared errors. Ideally, the ratio of the root mean sum of squared errors of the full model and the predictions should be close to one.

Statistical analysis of the data was done with R, with linear mixed model fitting with the lme4 and nlme packages and hypothesis testing with the lmerTest package.

Results

One novice’s data was unusable due to equipment failure. From 10 participants and 60 sutures, we extracted 1492 grasps, 475 from experts and 1017 from novices. For the pupil analysis, we were left with 448 grasps, 276 from novices and 172 from experts.

Number of grasps

Based on data extracted from surgical videos, we found that novices performed more grasps (mean ± SD) per suture than experts (18.03 ± 3.45, t = 5.22, P < 0.001) (Fig. 3A). We also found significant effects between surgeon skill and suture phase on the number of grasps. As shown in Figure 3B, large grasping differences were seen in the NH and Knot 3 instead of Knot 2. The effect of suture phase on the number of grasps (Fig. 3B) was significant F (3, 220.88) = 3.48, P = 0.017 as well as the effect of skill F (1, 7.80) = 9.37, P = 0.016, however, the interaction of phase and skill was not significant F (3, 220.88) = 1.57 P = 0.198.

The novices were significantly more likely to fail at grasping, at suture level the log-odds for failing to grasp = 1.53 (SD = 1.16), $\chi^2 = 8.04$, P = 0.005. On average, novices failed
29.58% and experts 15.20% of grasps in a suture. Suture phase also had a significant effect on the odds of failing to grasp $\chi^2 = 7.90, P = 0.048$ and the interaction effect of skill and phase, $\chi^2 = 20.45, P < 0.001$. Main effect of skill was not significant.

In the LOOC, the RMSE for the original fitted data at suture level was 13.00 and for the predicted data 13.60, or a ratio of 0.96.

Grasps and time to completion

The effect on suturing duration was modeled with the number of grasps, and the interaction of grasps and skill as fixed effects. Average suture completion time for novices was $175.2 \pm 73.2$ s and for experts $76.7 \pm 17.1$ s. Suturing duration as a function of number of grasps for experts and novices is shown in Figure 4.

The average intercept for the time to completion was $52.67 \pm 13.41$ s. For experts each grasp was associated with increase in task time of $1.52 \pm 0.67$ s, ($t = 2.27, P = 0.028$), and for novices of $3.45 \pm 0.71$ s. The difference to experts was statistically significant with $t = 2.704, P < 0.001$. The random intercept of participant SD was 16.65. The mean RMSE for the original model was 17.09 and for the predicted data 21.04, a ratio of 0.81.

The correlation between the grasps and suturing duration could be due to sutures where the novices struggled more. To test this, we fitted the model to a dataset where sutures with more than 30 grasps were excluded (leaving $n = 45$ sutures, 15 from novices). In these sutures, the average completion time for novices was $125.2 \pm 40.0$ s and for experts $76.7 \pm 17.1$ s. For novices each grasp was associated with increase in suturing duration of $1.57 \pm 0.86$ s, and for experts $1.44 \pm 0.54$ s, but the difference was not statistically significant ($t = 0.162, P = 0.872$).

Grasps and suturing performance

Among the three surgeons who evaluated the sutures using the UWOMSA knot-tying module, Kendall’s rank correlation coefficients for the three criteria were: quality of the knot $= 0.89$, efficiency $= 0.88$, handling $= 0.85$, indicating good level of correlation.

The effect of suturing performance on the number of grasps was modeled with UWOMSA efficiency-score (B), knot quality score (A) and skill as fixed effects. Skill was included to control overall differences between novices and experts. The handling score (C) was excluded because it was strongly correlated with the efficiency score.

One-point increase in the UWOMSA efficiency-score was associated with $5.30 \pm 1.38$ fewer grasps, $t = -3.84, P < 0.001$. The effect of the UWOMSA knot quality score was not significant. The LOOC mean RMSE for the original model was 10.39 and for the predicted data 11.95, or a ratio of 0.87. The number of grasps and UWOMSA scores are displayed in Figure 5.

Task-evoked pupil response during grasping

The effects of grasping on pupil dilation were modeled with grasping phase (pre-/post-grasp), skill and their interaction as fixed effects.

Task-evoked pupil response during grasping reveals different patterns between novices and experts. As shown in Figure 6, novices pupil size change (%) started to increase one second before the moment of grasping ($T = 0$ s) and continued to reach to the peak value about half second after grasping. In contrast, experts reduce their pupil size change (%) during the grasping.

When the pupil windows were averaged by skill, the effect of suture phase on minimum-to-maximum pupil difference was significant for skill, ($P < 0.001$) and for the interaction of phase and skill ($P = 0.002$), but not for phase alone ($P = 0.8494$). For mean pupil size change, the effect of skill was significant ($P < 0.001$) and phase ($P < 0.001$), and their interaction...
(P < 0.001). When the pupil windows were averaged by suture, the effects were not significant.

Discussion

The most significant results from this study were that novices performed significantly more grasps in the suturing task under microscope (Fig. 3), the increasing number of grasping was correlated with increased task time in novices (Fig. 4) and lowered performance score (Fig. 5) overall. These findings imply that analysis of surgeon’s grasping behavior can reveal details about their skill and performance.

The novices were significantly more likely to fail at grasping compared to experts (Fig. 3) This phenomenon was more prominent during the needle handling phase, indicating that novices had difficulty using the needle holders effectively under a microscope. Such a finding is consistent with what was reported by Martinec, and Zheng, where tool complexity and image incompatibility reduce the maneuverability of surgeons in their early phase of surgical training. The differences in grasp attempts were large enough that monitoring the number of grasps (including failed grasps) could reveal the trainee’s overall skill even from a few trials.

A direct result of the increasing number of grasps was the extension of task completion time. (Fig. 4). A study by Bann et al. indicated that the ratio of tool movements and completion time does not significantly differ between novices and expert surgeons. Our results, however, imply that the ratio of grasps and completion time can indeed differ based on skill. This effect was larger with sutures where the completion time was longer. One explanation could be that in novices, increase in grasps was linked to general lack of movement efficiency, whereas experts were still able to perform the other movements efficiently even if they occasionally required more grasp attempts. Analyzing the effect of grasps on completion time could thus be used to evaluate trainee’s overall efficiency.

In a study of surgical errors, Husslein et al. concluded that error counts could be used as a marker of surgical performance. We found that counts of grasping attempts correlated strongest with UWOMSA knot-tying module’s efficiency score (Fig. 5). The implication for surgical education is that we can assess surgeons’ performance according to the UWOMSA instrument’s criteria without expert judgment by simply monitoring the number of grasping attempts from surgical videos by computer vision methods.

In Figure 6, the pupil dilation coincides with the grasping moment, with statistically significant differences between novices and experts after pupil dilations. The result suggests that the novices and experts may have perceived the difficulty of grasping differently. This is in line with previous studies on task-evoked pupillary dilations, including those made in surgical tasks. However, the differences in peak pupil dilations were less than 1%. In experiments designed specifically to investigate task-evoked pupil dilations, the pupil size has typically increased between 2% and 12%. In microsurgery, the field of view is constantly changing, potentially affecting the amount of light reaching the surgeon’s eye, which consequently makes the real-time detection of task-evoked pupil dilation difficult. Assessing skill from grasp-related pupil dilations would require environments where the external factors can be controlled well.

Although our study was limited by the low number of participants, the number of trials for data collection was relatively high to protect statistical validity. The prediction and full model squared error comparisons also supported the validity of the results. The limited sample size was partially attributable to the time-consuming manual video annotation of grasps.
Another limitation related to the participants is the large skill gap between the novices and the experts. Some of the effects we observed may be seen only in participants that have very limited experience with microsurgical techniques. However, in our comparison of the number of grasps and UWOMSA scores, we saw that the grasps predicted performance even between expert surgeons. This supports the idea that analysis of grasps could reveal performance differences between more homogenous groups. Still, the capability of distinguishing finer performance differences (e.g., between intermediate and expert surgeons) should be investigated more explicitly in future studies.

The surgical assessments using UWOMSA were cross-disciplinary, where experts were plastic surgeons and evaluators were neurosurgeons, which may have some effects on surgical assessments. However, the assessment was relatively straightforward, all the assessments were done completely in randomized order and blinded to expertise level and using an objective rating scale. Nevertheless, the cross-disciplinary assessment may have some small effect on how some aspects were assessed and scored.

With this work, we have aimed to contribute to the ongoing research toward more objective surgical skill evaluation methods. Currently, objective surgical performance evaluation is based on assessment tools that consist of grading scales and checklists. Because such evaluation still requires expert surgeon’s manual input, some studies have investigated more automated data-based methods such as surgical tool motion analysis. These works have analyzed for example, surgeon’s motion economy, with the assumption that expert surgeons will display fewer excess movements. We believe that analyzing grasping attempts could be used similarly to aid objective evaluation of skill and performance in microsurgical training environments.

As surgeon’s performance is evidently affected by numerous other factors as well, we anticipate that grasp monitoring could be used to supplement information gained from the other tool-use metrics. Modern computer vision methods could be used both to track the tool movements and to identify the grasping attempts, thus providing us with more information about the surgeon’s performance. Since grasping with tools is a common action in surgery, it is likely that such analysis could be used in other surgical disciplines as well.

Conclusions

Our results indicate that grasping with surgical tools is challenging for novices in their early phase of performing a surgical task under a microscope. Strong and significant correlations were observed between the number of grasps and task time and suturing performance. Detecting the grasping events automatically from the videos recorded through the surgical microscope could enable objective evaluation of surgical skills during microsurgical performance.

Supplementary Materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jss.2022.09.018.